Taming the River of Data

New Software Tools Fuse Intelligence From Many Sources

By GLENN W. GOODMAN Jr.

With U.S. satellites, supercomputers, spies and soldiers gathering information day and night around the world, U.S. Army intelligence is drowning in data.

“While we always need more and better access to sources of information in support of answering the hard questions for the war fighters, the heart of our intelligence challenge is not collection,” said Maj. Gen. John Kimmons, who heads the Army’s Intelligence and Security Command at Fort Belvoir, Va. “It’s to fully leverage intelligence that we’ve already collected.”

To make sense of all that data, new software tools are being put to the test by Kimmons’ command’s Information Dominance Center (IDC) at Fort Belvoir, Va.

Kimmons’ command collects and processes some intelligence information, but its primary role, as part of the joint Defense Department intelligence team, is analyzing and fusing data from around the globe. The Information Dominance Center has gained unusual access to national intelligence data through hard-fought agreements with intelligence agencies such as the National Security Agency.

Yet, as Kimmons said, “Our analysts spend far too much time assembling and organizing data, as opposed to fusing it and analyzing it, to understand where it’s leading. Moreover, most of our analysts are working with fractions of all the data that’s already collected and reported.

“So how do they quickly integrate the latest ambiguous bits and pieces within the broader context of all the intelligence there is, for better, more complete understanding? Rapid all-source fusion analysis is what we are after — an intel question posed one time, and rubbed against the national, theater and tactical holdings that we already have to place it in a better context. And that requires information and database sharing, data structuring and the use of advanced software tools.”

So the IDC is pioneering the use of intelligence exploitation tools that help analysts “rapidly establish threat association and linkages, recognize threshold events, activity patterns and anomalies, understanding the significance of information ‘buried’ within large volumes of collected material,” said Lt. Gen. Keith Alexander, deputy chief of staff of the Army for intelligence (G-2).

Some software produces adaptable or dynamic signature graphs, which establish a baseline of normal behavior and sound an alert when thresholds are exceeded. For example, they might speak up when radios begin operating in areas where signals have never been transmitted before, when regular things stop happening, or when densities change or certain types of movements occur.

Kimmons compared it to a credit card company, whose anti-fraud software scans the purchases and transactions of cardholders to establish similar baseline behavior patterns. The purchase of a tuxedo in Hong Kong by a blue-jeans kind of guy from the United States may set off alarm bells, drawing a call from a company representative within minutes.

“Now, if they can do that with millions and millions of transactions and bits of data,” Kimmons said, “then why can’t we harness that power in the software sense to let us rake across terabytes of data rapidly and find the little pieces, the changes, the anomalies, the kind of things that a brigade or battalion commander wants to know — that certain threat information has been reported by somebody, somewhere, sometime, or the confluence of reporting has

come together to tell him in his sector of Baghdad that right now there is either an opportunity to do something or there's a target or a threat or a danger?"

There are a few critical requirements to making it all work, Kimmons said.

"You've got to get near-real-time access to all of the collected data — all of the humint [human intelligence], signals intelligence and imagery intelligence — from the PDA [personal digital assistant] of the squad on the street to things that are flying around above their heads, regardless of classification. You've got to structure the data in ways that let you rapidly search and visualize on your computer screen how the pieces interrelate and understand their significance and detect changes over time. Just like the cop in New York City does as he walks around his precinct and sees a new face or a new car and then talks to his informants."

Classifying the Data

For the military, such a system faces the extra challenge of classification. An important piece of data may have come from a source so secret that revealing the information may harm the source. Through tortuous negotiations with intelligence agencies around the government, the IDC has access to a network of databases that provide information at all sorts of classification levels, many with conditions on their use and protected by a myriad of code words.

But Kimmons envisions a system that would automatically scrub such information so that commanders with a clearance of secret or below could use it.

"The idea is that by taking this interrelated series of databases, we can come a lot closer, using mostly commercial advanced software tools and techniques, to determining the significance, in a timely way, of collected pieces of intelligence," he said.

But the key is to provide more than access to data.

"Access to a lot of databases is good, but it's also insufficient," Kimmons said. "Even if all the databases that exist in the U.S. intelligence community — hundreds and hundreds — were all linked together somehow, you would not be able to fully leverage them or understand the significances that pertain to your particular area of interest unless you have a process to inject all that information and organize and structure it in a way that lends itself to visualization on a common geospatial [background]. Our capability is not at 100 percent, but it validates our ability to do it.

"What holds us back is not the technological difficulty; it's the policy spider webs and reluctance to share intelligence at increasingly lower levels in near real time, where it is tactically relevant." •